

ARTICLE

Evaluation of the estimation and classification performance of NONMEM when applying mixture model for drug clearance

Ka Ho Hui | Tai Ning Lam

School of Pharmacy, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, Hong Kong

Correspondence

Tai Ning Lam, 8th floor, Lo Kwee-Seong Integrated Biomedical Sciences Building, Area 39, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong.
Email: teddylam@cuhk.edu.hk

Funding information

No funding was received for this work.

Abstract

Maximum likelihood estimation of parameters involving mixture model is known to have significant and specific patterns of errors. Population pharmacokinetic (PopPK) modeling using NONMEM is no exception. A few relevant studies on estimation and classification performance were done, but a comprehensive study was not yet available. The current study aims to evaluate performance and likelihood ratio test (LRT)-based true covariate detection rate when fitting a bimodal mixture of drug clearance (CL) in NONMEM. A large number of PopPK datasets with various settings were simulated and then estimated. The estimates were compared to the simulated values and summarized. The separation between the CL distributions of the two subpopulations is systematically overestimated. The major factor associated with the performance is the change in the minimum objective function value after removing the mixture component ($dOFV$). Other significant factors include estimated disparity index (DI), estimated mixing proportion, and number of subjects in the dataset. Small $dOFV$ and large estimated DI are associated with the worst performance. Omitting a true mixture resulted in reduced true covariate detection rates. It is recommended that on top of routinely generated standard errors and model diagnostics, $dOFV$, and other factors when necessary, should be taken into account for the evaluation of performance when fitting mixture model using NONMEM. In addition, when fitting mixture model for CL is intended, the mixture component should be introduced prior to LRT-based covariate model development for CL .

Study Highlights**WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?**

Estimation of a bimodal drug clearance (CL) distribution with mixture model in NONMEM using likelihood maximization algorithms is subject to biases related to the disparity.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *CPT: Pharmacometrics & Systems Pharmacology* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

WHAT QUESTION DID THIS STUDY ADDRESS?

The factors most associated with the estimation performance of parameter estimates, classification performance, and true covariate detection rates were identified and quantified.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

The change in objective function value after removing the mixture component (*dOFV*) is the major factor associated with estimation performance of *CL*-related parameters and classification performance. Other significant factors include estimated disparity index (*DI*), mixing proportion, etc. The performance is worsened when *dOFV* is small and estimated *DI* is large. These associations are quantified. Omitting a true mixture is associated with reduced likelihood ratio test-based true covariate detection rates.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?

dOFV, and other factors when necessary, should be taken into consideration when fitting and reporting population pharmacokinetic models with a mixture component. Mixture model should be evaluated for implementation before covariate model development where applicable.

INTRODUCTION

The mixture model functionality was first available in NONMEM IV, released in 1992.¹ On top of the routinely assumed continuous Gaussian distributions of random effects in between-subject variability model, mixture model assumes that the population consists of two or more subpopulations, each having its own model.^{2,3} The prevalence of reports of population pharmacokinetic (PK)/pharmacodynamic models with a mixture component estimated using NONMEM has increased over the past 2 decades. Mixture model is most commonly (28%) used to describe multimodal distributions of drug clearance (*CL*).^{4–33} A summary of these reports is available in Supplementary File S1.

Unfortunately, maximum likelihood estimation (MLE) of mixture model is known to have systematic patterns in estimation errors. Lourens et al. explained that the estimation errors are particularly concerning when the disparity between two distributions in a normal mixture is small (which is mainly driven by both small differences in means and large variances of the two distributions), where the disparity is more often overestimated than underestimated.³⁴ This issue with disparity is relevant to the estimation of multimodal distributions of PK parameters, hence corresponding population PK (PopPK) parameters defining such distributions. The estimation errors of individual PK parameters and individual classifications are also in doubt because they are conditioned on the estimated population parameters. Because the estimated PK parameters are essential determinants of dosage regimens, such patterns could compromise clinical dose decisions.

A few studies had done investigations related to mixture model in NONMEM.^{35–38} Among them, Carlsson et al., Kaila et al., and Yoon et al. focused on PopPK data and multimodal distributions of *CL*.^{35,37,38} However, these three studies separately looked into individual parameter estimates, individual classifications, and estimates for typical values of *CL*, respectively, under different study settings. Besides, they did not explore the estimation errors of non-*CL*-related parameters. More importantly, they fixed several factors that could influence estimation and classification errors. (Note that the word “factors” is used throughout this text to refer to all parameters and statistics that may associate with such errors.) In particular, all of them fixed the sampling schedule plus one or more of the followings: variances of parameters, the dominating subpopulation, the mixing proportion, number of subjects, and true values of non-*CL*-related parameters. The implication of fixing some of these population parameters remains unknown.

Additionally, no study has described the pattern of estimation errors when both mixture model and covariate effect are involved. Whether omitting a true mixture component would affect the rate of true covariate detection based on likelihood ratio test (LRT; the standard test for covariate model development in PopPK modeling) remains an open question.

This simulation study aims at narrowing the research gaps stipulated above by conducting more comprehensive and integrated investigations into the influences of various factors on the estimation of all parameters of interest. The following are the objectives of the study:

1. To simulate a large set of PopPK datasets without covariate effect and another set with covariate effect, using various parameter settings, assuming bimodal distributions of CL , and then perform estimation on the simulated datasets to generate the study databases.
2. To evaluate the performance of estimation and classification against various factors, including the disparity measures (namely $dOFV$, the change in objective function value [OFV] after removing the mixture component, and disparity index [DI], both elaborated later in the Methods section).
3. To compare the LRT-based true covariate detection rates with versus without the mixture component.

METHODS

Definitions – Pharmacokinetic parameters and symbols

The following PK parameters are defined: CL , volume of distribution (V_d), and absorption rate constant (k_a). The following modifications apply to these PK parameters:

- (population parameters) TV is the prefix for “typical value of,” (i.e., $TVCL$, TVV_d , and TVk_a);
- (population parameters) CV is the prefix for “coefficient of variance of” (i.e., $CVCL$, CVV_d , and CVk_a [defined as $\sqrt{e^{\omega^2} - 1} \times 100\%$, where ω^2 stands for the variance of the random effects of the respective PK parameters]); and
- (individual parameters) $caret$ (^) is the accent representing individual value (i.e., \widehat{CL} , \widehat{V}_d , and \widehat{k}_a).

σ is the population parameter quantifying residual unexplained variability (RUV), $\theta_{COV,CL}$ is the population parameter quantifying the covariate effect size, and γ is the skewness of the distribution of individual covariate values (defined as $\mu_3/\mu_2^{3/2}$, where μ_2 and μ_3 are the second and third central moments of the normal distribution, respectively).

By introducing a bimodal mixture for CL , extra parameters apart from the above are needed. The mixing proportion ($MIXP$; a population parameter) and the individual probability (P_{mix} ; an individual parameter) stands for the probability for a random subject and the probability for a specific subject with previous observation(s), respectively, of belonging to subpopulation 1 or 2 (as indicated by the subscript, i.e., $MIXP_1$ vs. $MIXP_2$ and $P_{mix,1}$ vs. $P_{mix,2}$). Note that $MIXP_1 + MIXP_2 = P_{mix,1} + P_{mix,2} = 100\%$. In addition, an extra degree of freedom is needed to differentiate the $TVCL$ s of the two subpopulations. For this purpose, the symbols $TVCL_1$ and $TVCL_2$ are defined such that

$TVCL_1 \leq TVCL_2$. In addition, rCL represents the ratio of $TVCL_2/TVCL_1$ (i.e., $rCL \geq 1$). $CVCL$ is assumed to be identical for the two subpopulations.

Besides, in this text, the term CL -related population parameters comprises $TVCL_1$, $TVCL_2$, $CVCL$, and $MIXP_1$, whereas the term CL -related parameters comprises the above plus \widehat{CL} and $P_{mix,1}$. The corresponding complementary sets, referred to as non- CL -related (population) parameters, consist of other (population) parameters defined in the previous paragraphs in this section that are being estimated.

Definitions – Terminology and symbols for estimation errors

In this text, the term estimation error refers to the difference of the estimated value from the true value. As widely understood, estimation error consists of two components, namely (1) bias, the systematic component, and (2) uncertainty, the random component.³⁹ In this text, bias is defined as the median of the estimation errors, denoted as $err_{rel,mdn}$ and $err_{abs,mdn}$ for relative and absolute errors, respectively. Meanwhile, uncertainty is represented by the ($\frac{1-x\%}{2}$ th percentile, $\frac{1+x\%}{2}$ th percentile) of the estimation errors, denoted as $err_{rel,x\%}$ and $err_{abs,x\%}$ for relative and absolute errors, respectively.

Definitions – Terminology and abbreviations for performance

In this text, the term estimation performance (EP) applies to all parameters to be estimated on the continuous scale. Good EP is indicated by the lack of biases and small uncertainty.

Each simulated subject is also classified to the subpopulation with higher P_{mix} . (For e.g., subpopulation 1 if $P_{mix,1} > P_{mix,2}$.) Based on the true and classified subpopulations of the subjects, the positive predictive values (PPV_1 and PPV_2), true positive rates (TPR_1 and TPR_2), and overall accuracy of correct subject classification ($\%CC$) were summarized for each dataset. The term classification performance (CP) applies to these summary values. Good CP is indicated by these values being close to 100%. The term performance, when used alone, refers to the collection of EP and CP.

Elaboration on disparity measure – $dOFV$

$dOFV$ is defined as the change in OFV after removing the mixture component. When the true disparity is large

(or small), introducing the mixture component often significantly improves (or does not much improve) the goodness-of-fit, resulting in a larger (or smaller) $dOFV$. Therefore, $dOFV$ can be regarded as a measure of the disparity of the bimodal CL distributions, especially for the true distribution. This is further illustrated in Figure 1.

Elaboration on disparity measure – DI

Lourens et al. mentioned the superiority of the overlapping coefficient (OVL) reported by Inman et al. over other statistics (see the separation index by Hosmer and the index D defined by Nityasuddhi et al.) in associating with EP of bimodal normal mixtures, where OVL refers to the overlapping area of the probability density plots of the two subpopulations.^{34,40–43} Lourens et al. defined DI as the complement of OVL (i.e., $DI = 1 - OVL$) as a measure of disparity.³⁴ DI is also applied in this study but with important modifications. Specifically, Lourens et al.

assumed $MIXP_1 = 50\%$ but cases where $MIXP_1 \neq 50\%$ are also considered in the definition of DI in this study; besides, Lourens et al. assumed the normal distribution, but this study uses the log-normal distribution, which better resembles the distribution of CL . Details of the definition and approximation procedure of DI in this study are presented in Supplementary File S2.

DI is dependent on the CL -related population parameters because these parameters characterize the probability density functions of CL . True DI and estimated DI , representing the true disparity and estimated disparity, refer to the DI computed based on the true and estimated values of CL -related population parameters, respectively. Therefore, theoretically, if estimated DI is significantly different from true DI , some forms of significant errors exist in the estimates for CL -related population parameters. As such, whereas DI is considered a measure of disparity, the discrepancies between true and estimated DI also reflect estimation errors. To facilitate understanding, DI is illustrated in shaded areas in light green in Figure 1.

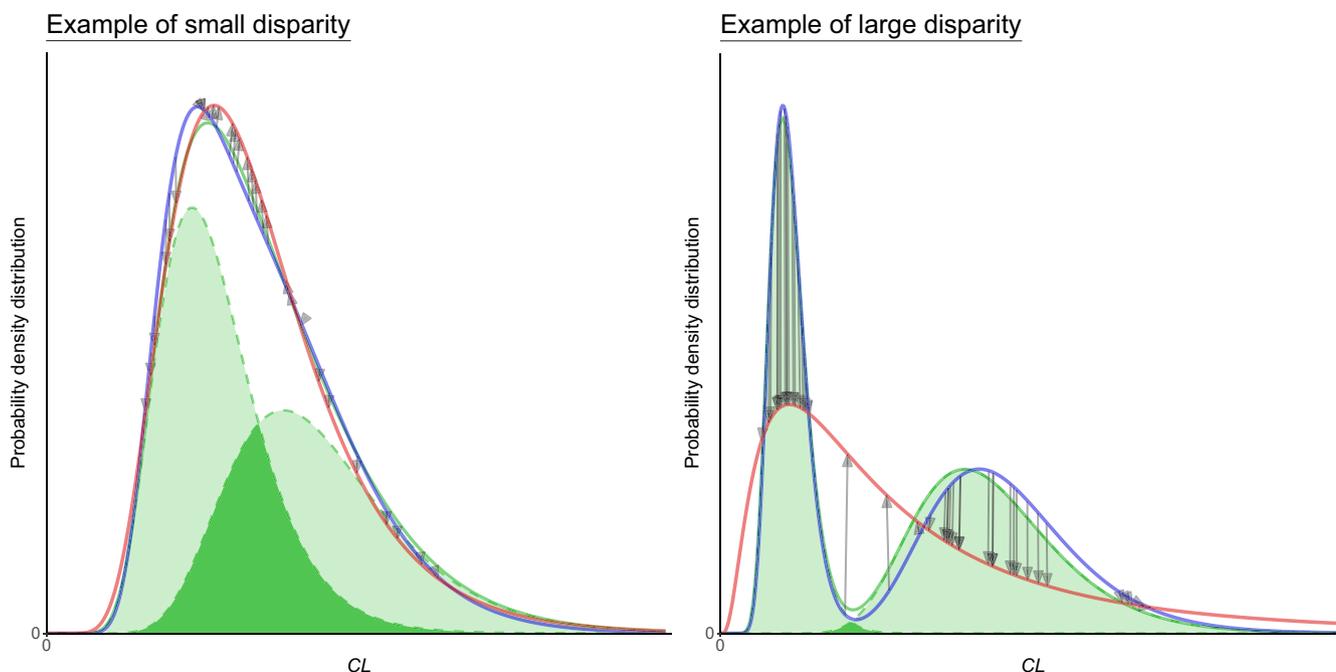


FIGURE 1 Elaboration of $dOFV$ and DI . A dataset was taken from the study database to exemplify the probability distribution of CL for each of small disparity (left) and large disparity (right). In each graph, the two dashed green lines represent the $MIXP$ -weighted probability density curves of the true (simulated) CL distributions of subpopulations 1 (left) and 2 (right). The areas under these lines are shaded in deep green and light green, which represent the overlapping area and distinct area (which defines the disparity index [DI]), respectively. It is visually intuitive that DI is positively related to the disparity of the two CL distributions. The solid green line represents the superimposed density of the two dashed green lines. The blue line represents the superimposed probability density of the CL distributions estimated with the mixture component, whereas the red line represents one estimated without the mixture component. Each of the grey arrows represents the change in a subject's estimated CL (along the x -axis) and the corresponding likelihood (along the y -axis) when the mixture component is removed (thus always pointing from the blue line to the red line). See Supplementary File S5, Part III for further explanation. CL , drug clearance; $dOFV$, change in the objective function value; $MIXP$, mixing proportion

Databases generation – PopPK model

For all datasets, a one-compartment structural model with first-order absorption and elimination was used to simulate all datasets. All PK parameters were assumed to be log-normally distributed. A bimodal mixture was assumed for CL . A proportional error model was used to describe RUV. For datasets with covariate effect, covariate effects were modeled with the commonly used exponential model.

Databases generation – Parameter sampling and simulation design

Each dataset was simulated based on a randomly sampled set of the following eight parameters: number of subjects (N), $TVCL_1$, rCL , $CVCL$, $MIXP_1$, TVV_d , TVk_a , and σ , plus two other parameters: (1) CVV_d and CVk_a for datasets without covariate effect or (2) $\theta_{COV,CL}$ and γ for datasets with covariate effect. For each of these 10 parameters, a range of values was decided. Each range was further divided into three strata. All combinations of these strata were sampled, thus resulting in $3^{10} = 59,049$ combinations. For each combination, three datasets without covariate effect and one dataset with covariate effect were simulated. For each dataset, each virtual subject was administered a single oral dose of 500 units, followed by 12 samples for drug concentration.

Databases generation – Parameter estimation

Each simulated dataset was then subjected to parameter estimation twice: the first time assuming the true model structure (denoted *Mix*) and the second time with the mixture component removed (denoted *NoMix*). Datasets with covariate effect were subject to parameter estimation for another two times: similar to *Mix* and *NoMix* but assuming no covariate effect (denoted as *MixNoCov* and *NoMixNoCov*, respectively).

Databases generation – Collection of results for analyses

For datasets without covariate effect, all results were summarized in two spreadsheets: (1) for population-level data, where each row represents a dataset, and (2) for individual-level data, where each row represents an individual from a dataset; the spreadsheets are collectively referred to as the database without covariate effect. The same applied to the database with covariate effect. These spreadsheets contain all data being analyzed later.

Databases generation – Process automation and software packages used

The generation of the databases was performed with SUSE Linux Enterprise Server 11. All processes were wrapped by R (version 3.4.3), through which Perl-speaks-NONMEM was configured and called to execute NONMEM 7.4 (Icon plc).^{2,44,45} First-order conditional estimation with interaction, which is an MLE algorithm, was used for parameter estimation. Deployed R packages included data.table (version 1.11.4), magrittr (version 1.5), and readr (version 1.1.1).⁴⁶⁻⁴⁸ Figure 2 shows the R-based workflow for the generation of the databases.

Databases generation – Further details

Due to space limitation, many details regarding database generation and the ready-to-run R source codes used to generate the databases are put in Supplementary File S3. Note that the choices of the PopPK model, parameter space, and estimation algorithm were based on their respective prevalences in the literature (*c.f.* Supplementary File S1, Table S1.2) such that it is more likely for future PopPK studies to find similar datasets in our study for reference.

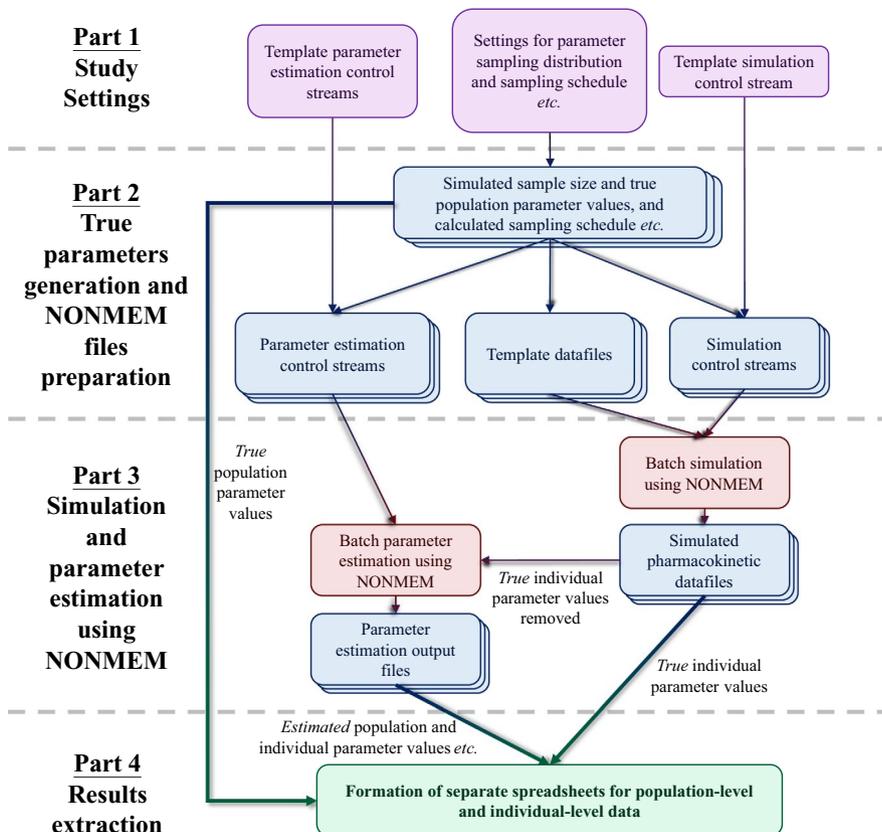
Data analyses – Datasets censoring

Only datasets whose estimation processes terminated with successful minimization and covariance step without any error message for both *Mix* and *NoMix* (plus *MixNoCov* and *NoMixNoCov* for analyses of covariate detection) were included in subsequent analyses. This is because parameter estimates tend to be erratic and unreliable otherwise.

Data analyses – Performance

Performance was first assessed graphically without stratification. Then, the analyses were repeated with single-factor stratification against N , all estimated population parameters, $\theta_{COV,CL}/rCL$ (which represents the covariate effect sizes relative to the subpopulation effect), $dOFV$, $dOFV/N$, estimated DI , γ , the individual change in OFV when the mixture component is removed ($d\widehat{OFV}$), the difference between subpopulation-based individual $OFVs$ ($d\widehat{OFV}_{grp}$), *etc.* To further investigate the effects of interactions among factors, multiple-factor stratified analyses were done against the significant single factors identified.

FIGURE 2 R-based workflow for the generation of the databases. The purple boxes represent settings that need to be specified before execution. The stacked blue boxes represent intermediate outputs for all datasets. The red boxes represent Perl-speaks-NONMEM and NONMEM processes. The green box represents the formation of the study databases



Data analyses – LRT-based true covariate detection rates

For each dataset with covariate effect, the changes in OFV after removing the covariate effect (ΔOFV ; which should be distinguished from $dOFV$) when fitting with versus without the mixture component (ΔOFV_{Mix} vs. ΔOFV_{NoMix} , respectively) were computed. LRT was then conducted at $\alpha = 0.05$ and 0.001 under the assumption that ΔOFV approximates the χ^2 -distribution with $df = 1$. The LRT-based true covariate detection rate based on ΔOFV_{Mix} and that based on ΔOFV_{NoMix} without stratification were computed. This was then repeated with simultaneous stratification by N and $\ln(\theta_{COV,CL})$. The detection rates based on ΔOFV_{Mix} versus ΔOFV_{NoMix} were then compared.

RESULTS

Database generation and dataset demographics

A database without covariate effect and another one with covariate effect consisting of 177,147 and 59,049 datasets (18,316,469 and 6,097,251 virtual subjects) were generated, where 73% and 42% of datasets were qualified for subsequent analyses, respectively. The generated databases are available in Supplementary File S4.

Estimated DI and $dOFV$ are positively correlated as expected, but the variability in estimated DI is much larger when $dOFV$ is small. See Supplementary File S5, Tables S5.1–2 and Figures S5.1–8 for detailed database demographics.

Unstratified analyses

Figure 3 and Supplementary File S5, Figure S5.9 show the results of unstratified analyses for the databases without and with covariate effect, respectively. Both databases display very similar results.

Biases are mostly mild for most parameters. However, there is a general trend to overestimate disparity (see Figure 3 and Figure S5.9, top-right graph). Despite mild biases for each CL -related population parameter, DI is systematically overestimated, where $err_{abs,mdn}$ can be as severe as +80% as true DI approaches 0%. $CVCL$ is increasingly systematically underestimated as true $CVCL$ increases. At true $CVCL$ greater than 45%, overall $err_{rel,mdn}$ is at -12% . The apparent biases of individual PK parameters at extreme values are explained in Supplementary File S5, section 5A.

Regarding uncertainty, the uncertainty of $MIXP_1$ is very severe, such that the $err_{abs,99\%}$ often cover 80% to nearly 100% of the whole range. Non- CL -related parameters generally have smaller uncertainty.

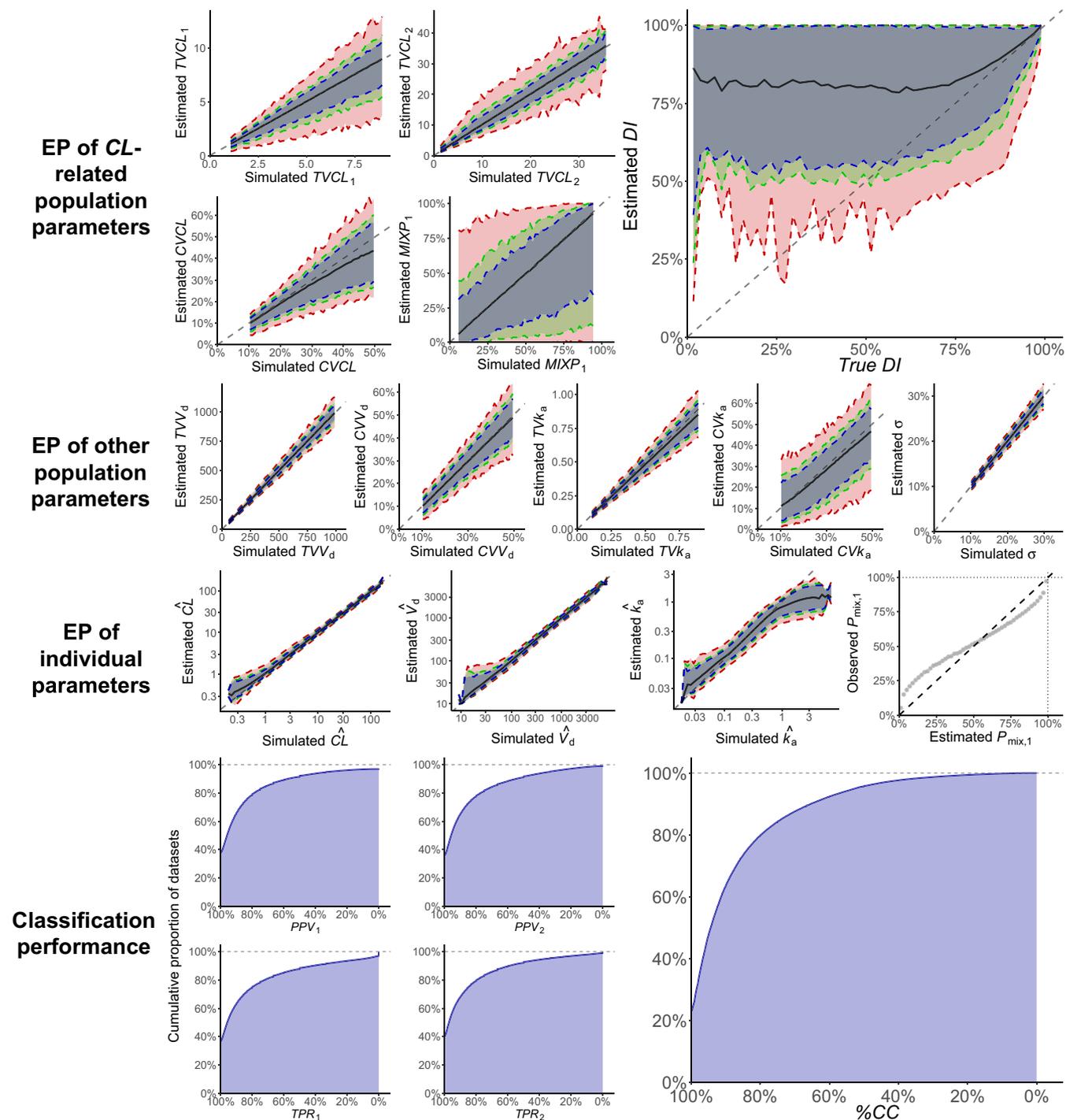


FIGURE 3 Unstratified analyses of estimation and classification performance for the database without covariate effect. This composite figure consists of four sections as indicated. There are three types of graphs in the figure (types I, II, and III). In each type I graph, the estimated values are plotted against the binned true values, where the dashed grey line is the line of identity, the solid black line is the median of estimates, and the shaded areas in blue, green, and red enclose the ranges of 5th–95th, 2.5th–97.5th, and 0.5th–99.5th percentiles of the estimates, respectively. The only type II graph is for $P_{mix,1}$. It has the observed $P_{mix,1}$ plotted against the binned estimated $P_{mix,1}$, where the dashed line represents the line of identity. Each type III graph is a cumulative polygon. The x-axes are arranged in decreasing order such that the area under the curve is positively related to classification performance. (See also Supplementary File S5, Part III.) *CL*, clearance; *CV*, prefix for “coefficient of variance of”; *DI*, disparity index; k_a , absorption rate constant; $MIXP_1$, mixing proportion for subpopulation 1; $P_{mix,1}$, individual probability of belonging to subpopulation 1; PPV_1 and PPV_2 , positive predictive value for subpopulations 1 and 2, respectively; TPR_1 and TPR_2 , true positive rate for subpopulations 1 and 2, respectively; *TV*, prefix for “typical value of”; $TVCL_1$ and $TVCL_2$, typical values of *CL* of subpopulations 1 and 2, respectively; V_d , volume of distribution; σ , population parameter quantifying residual unexplained variability; %CC, overall rate of correct classification. Caret (^), accented symbol for “individual value of”

$P_{\text{mix},1}$ shows a systematic trend to be overconfident regarding individual classification (i.e., overestimated when $>50\%$, and underestimated when $<50\%$), which is consistent with the tendency to overestimate disparity. For individual classification, around 80% of datasets can achieve $\geq 80\%$ in terms of %CC.

Single-factor stratified analyses

Figure 4 shows the results after single-factor stratification by $dOFV$ and estimated DI for $TVCL_1$, $TVCL_2$, $CVCL$, $MIXP_1$, $P_{\text{mix},1}$, and %CC for the database without covariate effect, whereas other results are available in Supplementary File S6. The results are very similar between both databases.

The most significant factors are found to be $dOFV$ and estimated DI . Larger $dOFV$ is associated with significantly better performance, especially in terms of reduced uncertainty. Notably, despite being highly correlated to $dOFV$, larger estimated DI is associated with performance differently. For example, when estimated DI increases beyond 95%, $err_{\text{rel},95\%}$ are (-49% , $+17\%$) and (-16% , $+36\%$) for $TVCL_1$ and $TVCL_2$, respectively, showing asymmetry consistent with the observed trend of overestimated disparity. Apart from $dOFV$ and estimated DI , the factors $MIXP_1$ and N behave similarly in that larger (sub)population sizes are associated with better performance, which agrees with basic statistical theory. Meanwhile, $dOFV$ and estimated DI associate minimally with EP of non- CL -related parameters. Regarding the estimation of $\theta_{\text{COV},CL}$, larger N and smaller estimated $CVCL$ (but not $dOFV/N$ nor estimated DI) display associations with better EP of $\theta_{\text{COV},CL}$.

Multiple-factor stratified analyses

Figure 5 shows the results after multi-factor stratification for $TVCL_1$, $TVCL_2$, $CVCL$, and $P_{\text{mix},1}$ for the database without covariate effect, whereas other results are available in Supplementary File S6. Both databases show very similar results when simultaneously stratified by $dOFV$, estimated DI , and estimated $MIXP_1$.

The multistratified analysis revealed the key finding of this study. Generally, when $dOFV$ becomes smaller and estimated DI becomes larger, EP of CL -related parameters and CP are worsened in the following patterns: $TVCL_1$ is negatively biased with negative uncertainty more severe than positive uncertainty; $TVCL_2$ behaves in the opposite direction of $TVCL_1$; $CVCL$ is negatively biased; $MIXP_1$ has very large uncertainty; $P_{\text{mix},1}$ shows overconfidence; and %CC is generally reduced.

Regarding EP of $\theta_{\text{COV},CL}$, biases of $\theta_{\text{COV},CL}$ remain inapparent even with multiple-factor stratification (see Supplementary File S6, last page). Large estimated $CVCL$

and small N are associated with more significant uncertainty of $\theta_{\text{COV},CL}$ estimates, which agrees with basic statistical theory that larger N and smaller variances are associated with better estimation accuracy.

Differences between fitting with versus without the mixture component regarding covariate effect

There were 15,571 datasets in the database with covariate effect that were included in subsequent analyses. Without stratification, the LRT-based true covariate detection rates were calculated to be 92.1% and 89.4% based on ΔOFV_{Mix} and $\Delta OFV_{\text{NoMix}}$, respectively. The stratified rates of LRT-based true covariate detection with *Mix* and *NoMix* are summarized in Table 1. (See Supplementary File S7 for the distributions of ΔOFV_{Mix} and $\Delta OFV_{\text{NoMix}}$.) Larger N and larger estimated $\theta_{\text{COV},CL}$ are associated with higher rates of covariate effect detection as expected. The results indicate that true covariate detection rates are consistently lower when the true mixture is omitted during estimation. By stratum, the largest observed absolute difference in the detection rates based on $\Delta OFV_{\text{NoMix}}$ versus ΔOFV_{Mix} is -29.5% . Besides, the differences in estimated $\ln(\theta_{\text{COV},CL})$ between *Mix* and *NoMix* have a mean value of -0.006 and a 2.5th–97.5th percentile range of (-0.118 to $+0.091$), demonstrating similar estimated $\theta_{\text{COV},CL}$ between *Mix* and *NoMix*.

DISCUSSION

Major findings

This study quantified, under the context of PopPK modeling and with the use of the MLE algorithm, the systematically overestimated disparity in estimating mixture model especially when the true disparity is small. The major factor associated with performance of mixture model is $dOFV$, followed by other significant factors, such as estimated DI . Datasets with small $dOFV$ and large estimated DI have more significant estimation errors. The associations between estimated DI versus performance are particularly obvious when $dOFV$ is not large (e.g., $dOFV < 30$). Meanwhile, the lack of association between $dOFV$ and estimated DI versus non- CL -related population parameters suggests that such phenomenon is likely specific to the parameters with a mixture distribution.

The estimation of $\theta_{\text{COV},CL}$ and that of mixture model do not appear to interact with each other. Besides, omitting the true mixture is associated with decreased true covariate detection rates especially when the result of LRT is neither extremely significant nor insignificant.

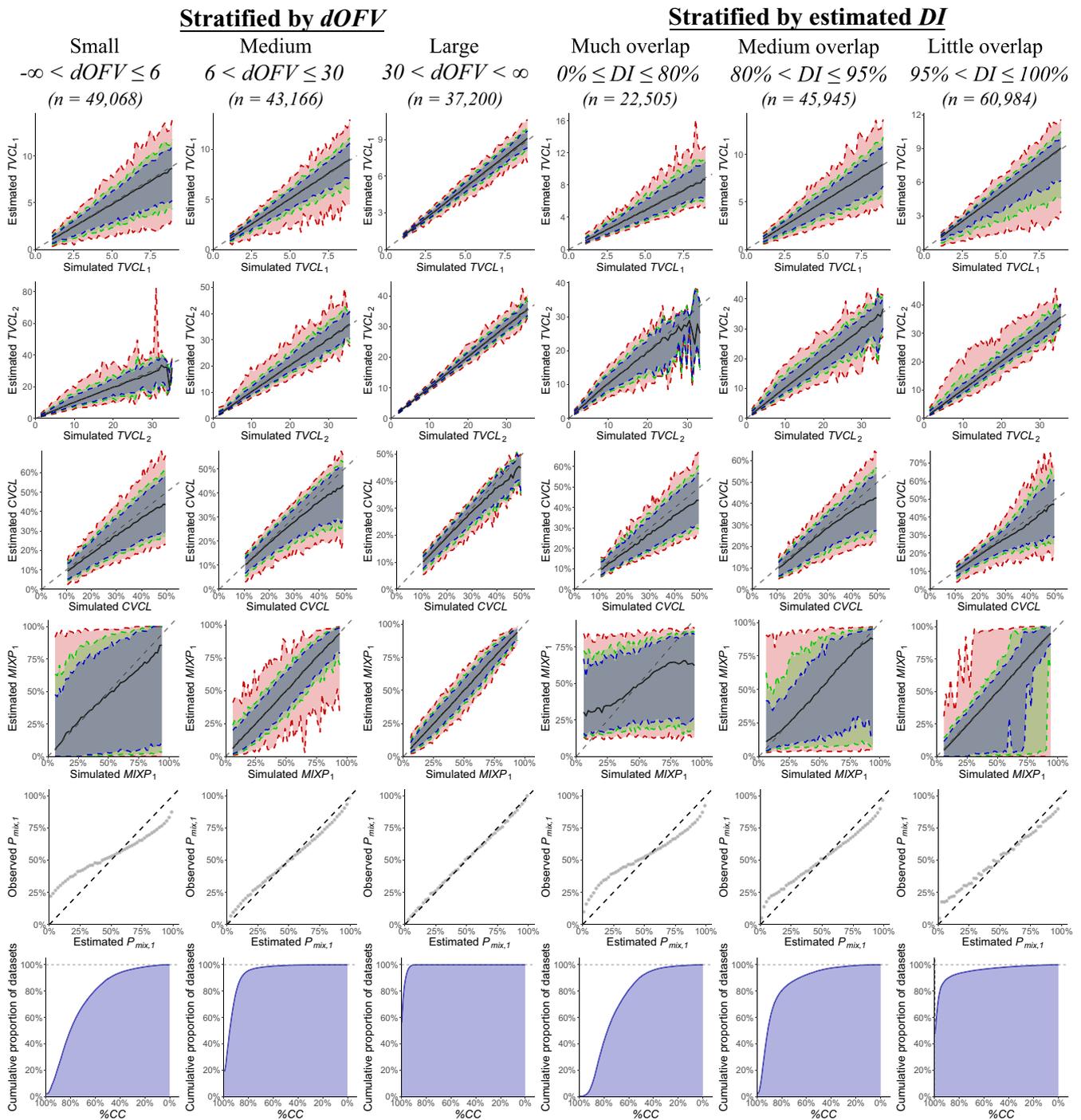


FIGURE 4 Stratified analyses for $TVCL_1$, $TVCL_2$, $CVCL$, $MIXP_1$, $P_{mix,1}$, and $\%CC$ using $dOFV$ and estimated DI as single factors for the database without covariate effect. In each column, the factor and range of stratification are shown at the header row of the figure. From the top to the bottom rows are the performance plots for $TVCL_1$, $TVCL_2$, $CVCL$, $MIXP_1$, $P_{mix,1}$, and $\%CC$, respectively. There are three types of graphs in the figure (types I, II, and III). In each type I graph, the estimated values are plotted against the binned true values, where the dashed grey line is the line of identity, the solid black line is the median of estimates, and the shaded areas in blue, green, and red enclose the ranges of 5th–95th, 2.5th–97.5th, and 0.5th–99.5th percentiles of the estimates, respectively. Each type II graph has the observed $P_{mix,1}$ plotted against the binned estimated $P_{mix,1}$, where the dashed line represents the line of identity. Each type III graph is a cumulative polygon. The x-axes are arranged in decreasing order such that the area under the curve is positively related to $\%CC$. (See also Supplementary File S5, Part III.) $CVCL$, coefficient of variance of drug clearance (CL); DI , disparity index; $dOFV$, change in objective function value after removing the mixture component; $MIXP_1$, mixing proportion for subpopulation 1; n , number of datasets in the stratum; $P_{mix,1}$, individual probability of belonging to subpopulation 1; $TVCL_1$ and $TVCL_2$, typical values of CL of subpopulations 1 and 2, respectively; $\%CC$, overall rate of correct classification

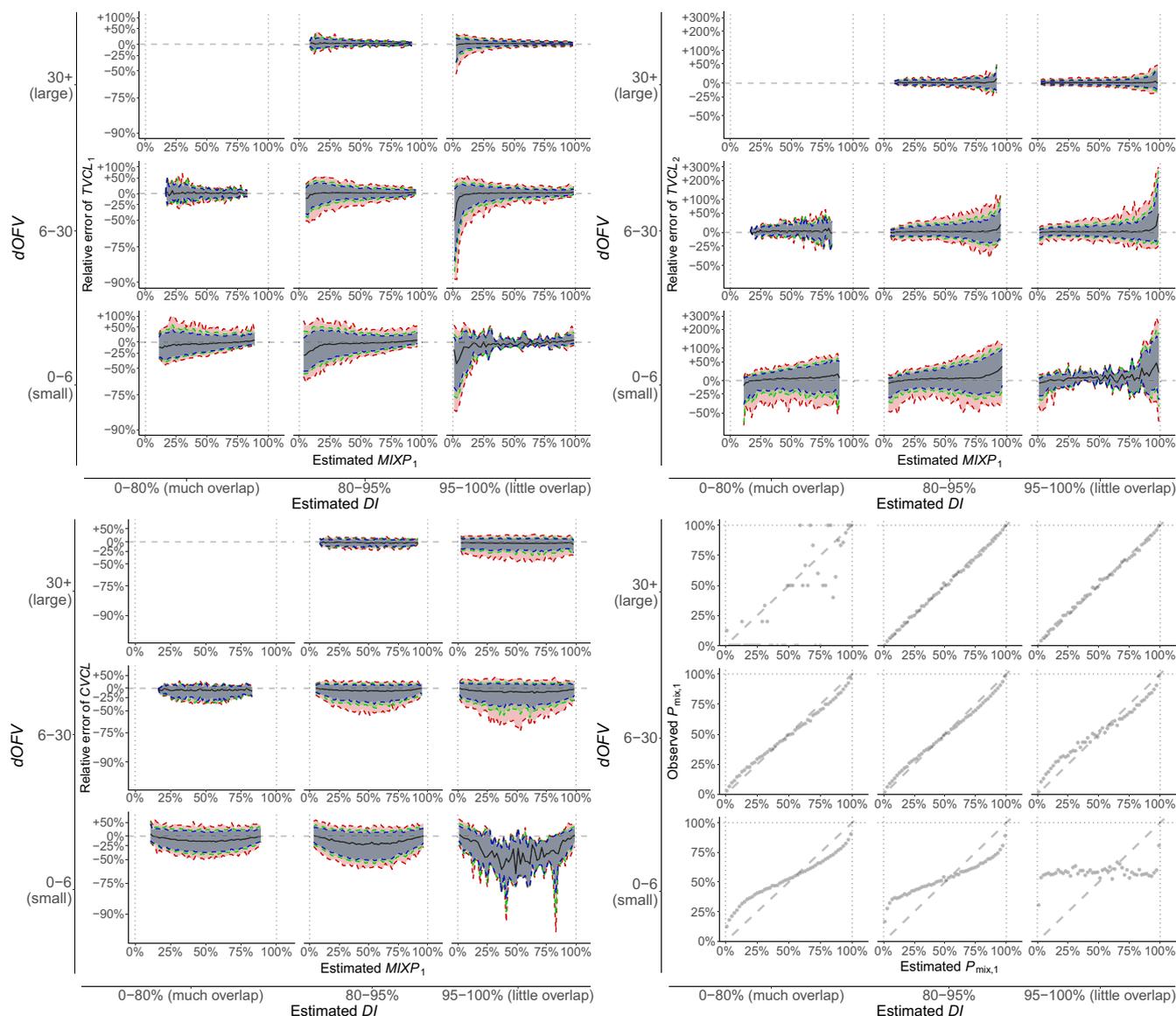


FIGURE 5 Stratified analyses for $TVCL_1$, $TVCL_2$, $CVCL$, and $P_{mix,1}$ against $dOFV$, estimated DI , and estimated $MIXP_1$ simultaneously for the database without covariate effect. This composite figure consists of four sections. The top-left, top-right, bottom-left, and bottom-right sections show the estimation performance for $TVCL_1$, $TVCL_2$, $CVCL$, and $P_{mix,1}$, respectively. In each section, the stratifications by $dOFV$ and estimated DI are represented by the outer y-axis and the outer x-axis, respectively. This results in a total of nine plots. In each plot, the x-axis represents estimated $MIXP_1$ (or estimated $P_{mix,1}$ for the bottom-right section) and the y-axis represents the relative estimation errors (or observed $P_{mix,1}$ for the bottom-right section). Except for the bottom-right section, in each plot, the dashed grey line is the line at 0% error, the solid black line is the median of relative errors of estimates, and the shaded areas in blue, green, and red enclose the ranges of 5th–95th, 2.5th–97.5th, and 0.5th–99.5th percentiles of the relative errors, respectively; besides, the inner y-axis is transformed to the scale of $\log(\text{estimated value}/\text{true value})$ such that the visual distance from 0% reflects the fold difference. For the bottom-right section, in each plot, the dashed grey line represents the line of identity. (See also Supplementary File S5, Part III.) $CVCL$, coefficient of variance of drug clearance (CL); DI , disparity index; $dOFV$, change in objective function value after removing the mixture component; $MIXP_1$, mixing proportion for subpopulation 1; $P_{mix,1}$, individual probability of belonging to subpopulation 1; $TVCL_1$ and $TVCL_2$, typical values of CL of subpopulations 1 and 2, respectively

Study limitations

Out of various model structures, this study has limited the exploration of mixture model estimation to a specific PopPK model and parameter space chosen based on its relative prevalence. For instance, the true model structure in this

study assumes identical $CVCL$ between the two subpopulations. Although this is not necessarily realistic, we noticed that estimating with separate $CVCL$ s is associated with frequent minimization or covariance step failure. In fact, most published PopPK models with mixtures of CL distributions assumed identical $CVCL$ (see Supplementary File S1, Table

TABLE 1 Stratified rates of LRT-based true covariate detection with versus without the mixture component

N	Estimated $\ln(\theta_{COV,CL})$				
	$-\infty$ to 0.12	0.12~0.20	0.20~0.32	0.32~0.43	0.43 to ∞
LRT-based true covariate detection rates at $\alpha = 0.05$					
200~300	98.2% (0.0%) [503]	100.0% (0.0%) [564]	100.0% (0.0%) [458]	100.0% (0.0%) [580]	100.0% (0.0%) [273]
100~200	90.5% (0.2%) [514]	99.8% (0.2%) [559]	100.0% (0.0%) [538]	100.0% (0.0%) [543]	100.0% (0.0%) [371]
75~100	77.0% (−1.8%) [444]	99.0% (−3.7%) [599]	100.0% (−0.3%) [605]	100.0% (0.0%) [536]	100.0% (0.0%) [481]
50~75	56.0% (−1.5%) [455]	96.5% (−6.8%) [542]	99.5% (−1.4%) [576]	100.0% (−0.2%) [524]	100.0% (0.0%) [619]
35~50	42.0% (−4.4%) [459]	90.0% (−13.3%) [510]	98.5% (−7.3%) [606]	99.8% (−0.7%) [447]	99.9% (0.1%) [669]
20~35	25.0% (−3.8%) [452]	72.7% (−13.5%) [466]	94.7% (−17.2%) [563]	98.3% (−8.3%) [411]	99.9% (−0.6%) [704]
LRT-based true covariate detection rates at $\alpha = 0.001$					
200~300	84.3% (0.2%) [503]	99.8% (−0.4%) [564]	100.0% (−0.2%) [458]	100.0% (0.0%) [580]	100.0% (0.0%) [273]
100~200	62.3% (−4.5%) [514]	93.0% (−7.2%) [559]	99.4% (−2.6%) [538]	100.0% (−0.2%) [543]	100.0% (0.0%) [371]
75~100	40.5% (−7.0%) [444]	75.8% (−16.7%) [599]	97.5% (−11.4%) [605]	100.0% (−1.9%) [536]	100.0% (0.0%) [481]
50~75	25.7% (−6.4%) [455]	60.3% (−17.7%) [542]	91.5% (−16.3%) [576]	99.8% (−4.4%) [524]	100.0% (−0.2%) [619]
35~50	17.0% (−8.3%) [459]	49.8% (−24.9%) [510]	77.2% (−25.9%) [606]	96.2% (−11.6%) [447]	99.7% (−0.6%) [669]
20~35	6.4% (−3.3%) [452]	32.2% (−20.2%) [466]	64.3% (−29.5%) [563]	86.6% (−17.5%) [411]	96.4% (−4.8%) [704]
LRT-based p values					
200~300	0.0059 (0.0038)	<0.001 (<0.001)	<0.001 (<0.001)	<0.001 (<0.001)	<0.001 (<0.001)
100~200	0.0208 (0.0208)	<0.001 (<0.001)	<0.001 (<0.001)	<0.001 (<0.001)	<0.001 (<0.001)
75~100	0.0658 (0.0671)	0.0044 (0.0106)	<0.001 (0.0012)	<0.001 (<0.001)	<0.001 (<0.001)
50~75	0.1424 (0.1346)	0.0083 (0.0248)	0.0026 (0.0050)	<0.001 (<0.001)	<0.001 (<0.001)
35~50	0.2038 (0.2044)	0.0215 (0.0517)	0.0047 (0.0178)	<0.001 (0.0029)	0.0015 (<0.001)
20~35	0.3321 (0.3215)	0.0473 (0.1073)	0.0100 (0.0544)	0.0031 (0.0166)	<0.001 (0.0013)

Note: Datasets are stratified by N and $\theta_{COV,CL}$, with the ranges specified on the leftmost column and the header rows, respectively. Each data cell represents a stratum of datasets. The table is divided into the top, middle, and bottom parts. For the top and middle parts, in each cell, the percentage on the left represents the proportion of datasets with the true covariate detected based on LRT conducted with the mixture component (at $\alpha = 0.05$ (top part of table) and 0.001 (middle part of table) with $df = 1$, equivalent to an increase in $OFV > 3.841$ and 10.828 after removing the covariate from the model, respectively); the percentage within the brackets represents the difference in the proportion when LRT was conducted without the mixture component, compared to LRT conducted with the mixture component (bolded cell if the difference is over 5%); the number within the square brackets represents the number of datasets within the stratum. For the bottom part of the table, in each cell, the number on the left is the mean value of LRT-based p values with the mixture component, whereas the number within the brackets is that without the mixture component. LRT, likelihood ratio test; N , number of subjects in the dataset; OFV , objective function value; α , significance level of the LRT; $\theta_{COV,CL}$, population parameter quantifying the covariate effect size.

S1.2). To maintain the quality of data analyses, the investigation of separate $CVCLs$ is excluded from the current study.

Besides, a factor that can influence $dOFV$ but is not accounted for in this study is the sampling schedule. Designing the number of samples and the timings of sampling can render numerous situations, thus requiring another large-scale study to analyze under the context of mixture model. Preliminary analyses also showed that sparse data (6 samples or less per subject) lead to frequent optimization failure under the current PopPK model. Therefore, the sampling schedule was fixed at 12 samples per subject in this study. (The mean is an average of 9 samples per subject for studies listed in Supplementary File S1 and Table S1.2.)

In addition, one may argue that investigating parameter estimates themselves (instead of the true values) or their derivatives as factors is not legitimate. However, it

should be emphasized that true values are unavailable in actual studies. To enable assessments of performance, it was decided that the estimated values instead of the true values are used as factors.

Insights for future PopPK studies based on the current results

Although the limited scope of this study implies that the reported numbers are only applicable to the specific model structure and study design, it is presumable that the trend of overestimated disparity extends to other settings in PopPK modeling. Therefore, based on present findings, we recommend that $dOFV$ should be considered in the evaluation of performance in PopPK mixture model fitting, on top of routinely obtained model diagnostics;

and when $dOFV$ is small, estimated DI and other factors can be regarded as auxiliary factors to assist the evaluation. Meanwhile, the database generated from this study acts as useful numerical reference regarding performance, especially when the model structure and study design are similar to this study. (An Excel-based database query interface is available in Supplementary File S8, which also features an automated calculator for the approximation of estimated DI .) Besides, mixture model should be evaluated for implementation before performing LRT-based covariate model development to mitigate the risk of not detecting true covariates.

Performance of routinely reported model diagnostics

The asymmetric standard errors (SEs) in NONMEM are often not good descriptors of uncertainty, especially in case of complex models.⁴⁹ Current results imply that in the context of mixture model, SE is unlikely representative of uncertainty as well. For example, Figure 6 illustrates the relative incapability of SEs for $TVCL_1$, $TVCL_2$,

$CVCL$, and $MIXP_1$ to capture the sizes and distributions of estimation errors when compared to non- CL -related population parameters. As such, bootstrapping might be more preferable to inform about uncertainty. The performances of these model diagnostics are of interest but outside the scope of this study. Further studies in this aspect are required.

Clinical relevance

CL is the determinant for drug exposure and thus an essential reference for dose decisions. Dose decisions based on PopPK studies are in demand everywhere, from preclinical studies to postmarketing drug use optimization studies and therapeutic drug monitoring. The misleading parameter estimates obtained from fitting mixture model in NONMEM can lead to the decisions of suboptimal or even ineffective or toxic doses. In fact, the patterns of estimation errors identified in different parameters have their respective relevances to the dose decision process. For instance, the biases and asymmetric uncertainty of estimated $TVCL_1$ and $TVCL_2$ can lead to suboptimal dose decisions. In addition,

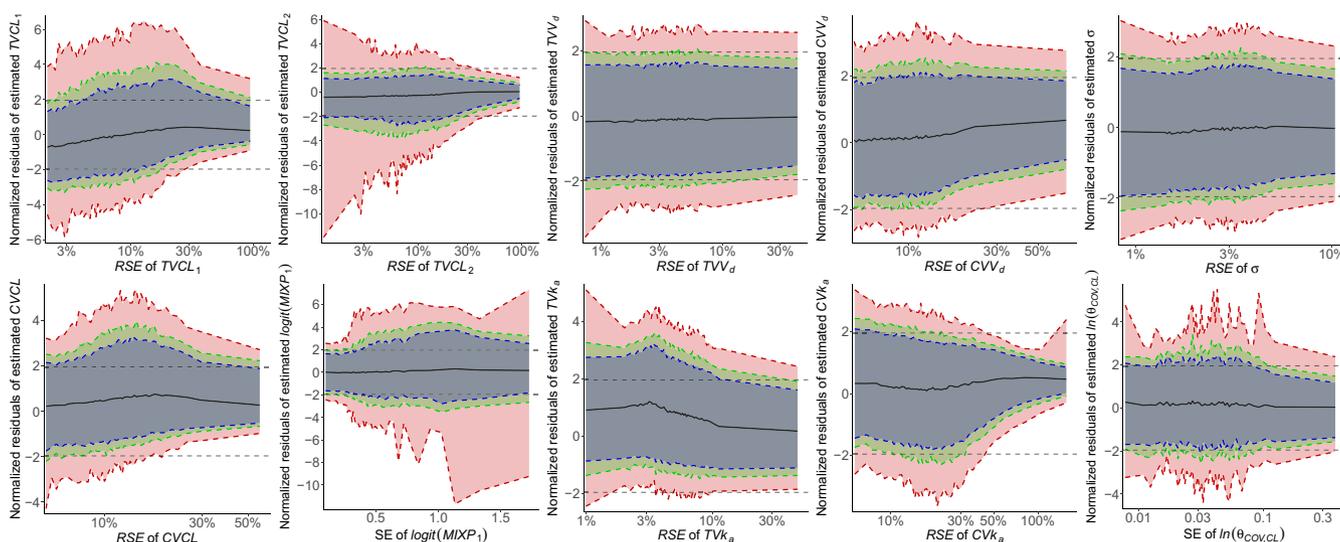


FIGURE 6 Distribution plots of normalized residuals of population parameter estimates versus standard errors reported from NONMEM. The 10 plots for $TVCL_1$, $TVCL_2$, $CVCL$, $MIXP_1$, TVV_d , CVV_d , TVk_a , CVk_a , and σ from the database without covariate effect, and $\theta_{COV,CL}$ from the database with covariate effect are shown in the corresponding plots in the figure, as indicated. For each of these population parameters, the normalized residuals are calculated as the standard z-scores of the sampling distribution (defined using the estimated values as means and the corresponding standard errors (SEs) or relative SEs ($RSEs$) as SDs , assuming a normal distribution for $\logit(MIXP_1)$ and $\ln(\theta_{COV,CL})$ or a log-normal distribution for other population parameters) at the true values. The normalized residuals are plotted against the binned SEs or $RSEs$, where the dashed grey horizontal lines enclose the central 95% range of a standard normal distribution (i.e., -1.96 to 1.96), the solid black line is the median of normalized residuals, and the shaded areas in blue, green, and red enclose the ranges of 5th–95th, 2.5th–97.5th, and 0.5th–99.5th percentiles (i.e., 90%, 95%, and 99% ranges) of the normalized residuals, respectively. Ideally, the green area should have sizes and positions that match the area enclosed by the dashed grey horizontal lines. CL , clearance; CV , prefix for “coefficient of variance of”; k_a , absorption rate constant; $MIXP_1$, mixing proportion for subpopulation 1; TV , prefix for “typical value of”; $TVCL_1$ and $TVCL_2$, typical values of CL of the subpopulations 1 and 2, respectively; V_d , volume of distribution; σ , population parameter quantifying residual unexplained variability; $\theta_{COV,CL}$, population parameter quantifying the covariate effect size

negatively biased *CVCL* can lead to overconfidence in the proportion of the target population that can benefit from the decided dose. In response, findings in this study can help numerically inform the average distributions of estimation errors and hence the optimality of dose decisions.

Meanwhile, omitting true covariates has the consequence of obtaining less accurate estimates of individual PK parameters. For this reason, the decision to not implementing mixture model before LRT-based covariate model development should be carefully considered.

CONCLUSIONS

The EP of *CL*-related parameters and CP when fitting bimodal distributions of *CL* in NONMEM is generally worse when *dOFV* is small. When *dOFV* is small, estimation errors are the most severe when estimated *DI* is large. Other relatively significant factors include estimated *MIXP*₁ and *N*. Modelers are recommended to take at least *dOFV*, and other factors when necessary, into account for the evaluation of performance, on top of results obtained from routinely done model diagnostics.

The estimation of covariate effect size and the estimation of mixture model do not significantly associate with each other. Omitting the mixture component when there exists a true mixture distribution would lead to decreased rates of LRT-based true covariate detection in covariate model development. Therefore, evaluation for implementing mixture model before covariate model development is preferable.

CONFLICT OF INTEREST

The authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS

H.K.H. and L.T.N. wrote the manuscript. H.K.H. and L.T.N. designed the research. H.K.H. and L.T.N. performed the research. H.K.H. analyzed the data.

REFERENCES

- ICON plc. History of NONMEM Development. https://nonmem.iconplc.com/nonmem_history. Accessed May 15, 2019.
- Boeckmann AJ, Sheiner LB, Beal SL. NONMEM Users Guides. Accessed, 2009.
- McLachlan GJ, Peel D. *Finite mixture models*. John Wiley & Sons, Inc.; 2004.
- Allegaert K, Holford NHG, Anderson BJ, et al. Tramadol and O-desmethyl tramadol clearance maturation and disposition in humans: a pooled pharmacokinetic study. *Clin Pharmacokinet*. 2015;54:167-178.
- Aruldas BW, Høglund RM, Ranjalkar J, et al. Optimization of dosing regimens of isoniazid and rifampicin in children with tuberculosis in India. *Br J Clin Pharmacol*. 2019;85:644-654.
- Barrett JS, Mondick JT, Narayan M, Vijayakumar K, Vijayakumar S. Integration of modeling and simulation into hospital-based decision support systems guiding pediatric pharmacotherapy. *BMC Med Inform Decis Mak*. 2008;8:6.
- Bieniczak A, Cook A, Wiesner L, et al. The impact of genetic polymorphisms on the pharmacokinetics of efavirenz in African children. *Br J Clin Pharmacol*. 2016;82:185-198.
- Brvar N, Mateović-Rojnik T, Grabnar I. Population pharmacokinetic modelling of Tramadol using inverse Gaussian function for the assessment of drug absorption from prolonged and immediate release formulations. *Int J Pharm*. 2014;473:170-178.
- Chen LF, Chen R, Huang SP, et al. Population pharmacokinetics of isoniazid in Chinese tuberculosis patients. [Chinese] *Chin Pharmaceut J*. 2017;52:2185-2191.
- Chirehwa MT, McIlleron H, Wiesner L, et al. Effect of efavirenz-based antiretroviral therapy and high-dose rifampicin on the pharmacokinetics of isoniazid and acetyl-isoniazid. *J Antimicrob Chemother*. 2018;74:139-148.
- Choi HY, Kim YH, Hong D, Kim SS, Bae KS, Lim HS. Therapeutic dosage assessment based on population pharmacokinetics of a novel single-dose transdermal donepezil patch in healthy volunteers. *Eur J Clin Pharmacol*. 2015;71:967-977.
- Dombrowsky E, Jayaraman B, Narayan M, Barrett JS. Evaluating performance of a decision support system to improve methotrexate pharmacotherapy in children and young adults with cancer. *Ther Drug Monit*. 2011;33:99-107.
- Dooley KE, Denti P, Martinson N, et al. Pharmacokinetics of efavirenz and treatment of HIV-1 among pregnant women with and without tuberculosis coinfection. *J Infect Dis*. 2015;211:197-205.
- Facca B, Frame B, Triesenberg S. Population pharmacokinetics of ceftizoxime administered by continuous infusion in clinically ill adult patients. *Antimicrob Agents Chemother*. 1998;42:1783-1787.
- Feng Y, Pollock BG, Coley K, et al. Population pharmacokinetic analysis for risperidone using highly sparse sampling measurements from the CATIE study. *Br J Clin Pharmacol*. 2008;66:629-639.
- Francis J, Zvada SP, Denti P, et al. A population pharmacokinetic analysis shows that arylacetamide deacetylase (AADAC) gene polymorphism and HIV infection affect the exposure of rifapentine. *Antimicrob Agents Chemother*. 2019;63:e01964-18.
- Garrido MJ, Habre W, Rombout F, Trocóniz IF. Population pharmacokinetic/pharmacodynamic modelling of the analgesic effects of tramadol in pediatrics. *Pharm Res*. 2006;23:2014-2023.
- Graves NM, Brundage RC, Wen Y, et al. Population pharmacokinetics of carbamazepine in adults with epilepsy. *Pharmacotherapy*. 1998;18:273-281.
- Green B, Crauwels H, Kakuda TN, Vanveggel S, Brochot A. Evaluation of concomitant antiretrovirals and CYP2C9/CYP2C19 polymorphisms on the pharmacokinetics of etravirine. *Clin Pharmacokinet*. 2017;56:525-536.
- Homkham N, Cressey TR, Bouazza N, et al. Efavirenz concentrations and probability of HIV replication in children. *Pediatr Infect Dis J*. 2015;34:1214-1217.
- Hussein R, Charles BG, Morris RG, Rasiah RL. Population pharmacokinetics of perhexiline from very sparse, routine monitoring data. *Ther Drug Monit*. 2001;23:636-643.
- Morris D, Podolski J, Kirsch A, Wiehle R, Fleckenstein L. Population pharmacokinetics of telapristone (CDB-4124) and

- its active monodemethylated metabolite CDB-4453, with a mixture model for total clearance. *AAPS J.* 2011;13:665-673.
23. Salman S, Sy SKB, Ilett KF, Page-Sharp M, Paech MJ. Population pharmacokinetic modeling of tramadol and its O-desmethyl metabolite in plasma and breast milk. *Eur J Clin Pharmacol.* 2011;67:899-908.
 24. Sherwin CMT, Saldaña SN, Bies RR, Aman MG, Vinks AA. Population pharmacokinetic modeling of risperidone and 9-hydroxyrisperidone to estimate CYP2D6 subpopulations in children and adolescents. *Ther Drug Monit.* 2012;34:535-544.
 25. Son H, Lee D, Lim LA, Jang SB, Roh H, Park K. Development of a pharmacokinetic interaction model for co-administration of simvastatin and amlodipine. *Drug Metab Pharmacokinet.* 2014;29:120-128.
 26. Stringer F, Ploeger BA, DeJongh J, et al. Evaluation of the impact of UGT polymorphism on the pharmacokinetics and pharmacodynamics of the novel PPAR agonist sipoglitazar. *J Clin Pharmacol.* 2013;53:256-263.
 27. Svensson E, van der Walt JS, Barnes KI, et al. Integration of data from multiple sources for simultaneous modelling analysis: Experience from nevirapine population pharmacokinetics. *Br J Clin Pharmacol.* 2012;74:465-476.
 28. Tanigawa T, Heinig R, Kuroki Y, Higuchi S. Evaluation of interethnic differences in repinotan pharmacokinetics by using population approach. *Drug Metab Pharmacokinet.* 2006;21:61-69.
 29. Vermeulen A, Piotrovsky V, Ludwig EA. Population pharmacokinetics of risperidone and 9-hydroxyrisperidone in patients with acute episodes associated with bipolar I disorder. *J Pharmacokinet Pharmacodyn.* 2007;34:183-206.
 30. Wilkins JJ, Langdon G, McIlleron H, Pillai G, Smith PJ, Simonsson USH. Variability in the population pharmacokinetics of isoniazid in South African tuberculosis patients. *Br J Clin Pharmacol.* 2011;72:51-62.
 31. Yang J, Ma P, Bullman J, Nicholls A, Chen C. Adjustment of the area under the concentration curve by terminal rate constant for bioequivalence assessment in a parallel-group study of lamotrigine. *Br J Clin Pharmacol.* 2019;85:563-569.
 32. Zhu M, Wu B, Brandl C, et al. Blinatumomab, a bispecific T-cell engager (BiTE®) for CD-19 targeted cancer immunotherapy: clinical pharmacology and its implications. *Clin Pharmacokinet.* 2016;55:1271-1288.
 33. Zvada SP, Denti P, Donald PR, et al. Population pharmacokinetics of rifampicin, pyrazinamide and isoniazid in children with tuberculosis: In silico evaluation of currently recommended doses. *J Antimicrob Chemother.* 2014;69:1339-1349.
 34. Lourens S, Zhang Y, Long JD, Paulsen JS. Bias in estimation of a mixture of normal distributions. *J Biometric Biostat.* 2013;4:1000179.
 35. Carlsson KC, Savić RM, Hooker AC, Karlsson MO. Modeling subpopulations with the \$MIXTURE subroutine in NONMEM: Finding the individual probability of belonging to a subpopulation for the use in model analysis and improved decision making. *AAPS J.* 2009;11:148-154.
 36. Frame B, Miller R, Lalonde RL. Evaluation of mixture modeling with count data using NONMEM. *J Pharmacokinet Pharmacodyn.* 2003;30:167-183.
 37. Kaila N, Straka RJ, Brundage RC. Mixture models and subpopulation classification: a pharmacokinetic simulation study and application to metoprolol CYP2D6 phenotype. *J Pharmacokinet Pharmacodyn.* 2007;34:141-156.
 38. Yoon S, Lim HS. Performance of a mixture model by the degree of a missing categorical covariate when estimating clearance in NONMEM. *Translat Clin Pharmacol.* 2019;27:141-148.
 39. Sheiner LB, Beal SL. Some suggestions for measuring predictive performance. *J Pharmacokinet Biopharm.* 1981;9:503-512.
 40. Hosmer DW Jr. On mle of the parameters of a mixture of two normal distributions when the sample size is small. *Commun Stat.* 1973;1:217-227.
 41. Inman HF, Bradley EL Jr. The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Commun Stat Theory Meth.* 1989;18:3851-3874.
 42. Nityasuddhi D, Böhning D. Asymptotic properties of the EM algorithm estimate for normal mixture models with component specific variances. *Comput Stat Data Anal.* 2003;41:591-601.
 43. Hosmer DW Jr. A comparison of iterative maximum likelihood estimates of the parameters of a mixture of two normal distributions under three different types of sample. *Biometrics.* 1973;29:761-770.
 44. R Core Team. R: A language and environment for statistical computing (2020). <https://www.R-project.org/>.
 45. Lindbom L, Pihlgren P, Jonsson E. PsN-Toolkit—a collection of computer intensive statistical methods for non-linear mixed effect modeling using NONMEM. *Comput Methods Programs Biomed.* 2005;79:241-257.
 46. Dowle M, Srinivasan A. data.table: Extension of 'data.frame' (2019). <https://CRAN.R-project.org/package=data.table>.
 47. Bache SM, Wickham H. magrittr: A forward-pipe operator for R (2014). <https://CRAN.R-project.org/package=magrittr>.
 48. Wickham H, Hester J, Francois R. readr: Read Rectangular Text Data (2018). <https://CRAN.R-project.org/package=readr>.
 49. Mishina EV, Booth BP, Gobburu JV. Comparison of confidence intervals derived using asymptotic standard errors and bootstrapping. *Clin Pharmacol Ther.* 2003;73:P24.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Hui KH, Lam TN.

Evaluation of the estimation and classification performance of NONMEM when applying mixture model for drug clearance. *CPT Pharmacometrics Syst Pharmacol.* 2021;10:1564–1577. doi:[10.1002/psp4.12726](https://doi.org/10.1002/psp4.12726)